Dental Issues in Youth across the Globe

Exploring risk factors of oral disease in children

Silas Lee
Springboard

Overview

What is a good indicator for dental health across the globe? Which countries and areas have the largest amount of oral disease in children under 12 and what is most correlated?

According to the World Health Organization, the most common oral diseases are dental cavities, periodontal (gum) disease, oral cancer, oral infectious diseases, trauma from injuries, and hereditary lesions. They also state that 60–90% of children and nearly 100% of adults have dental cavities across the globe.

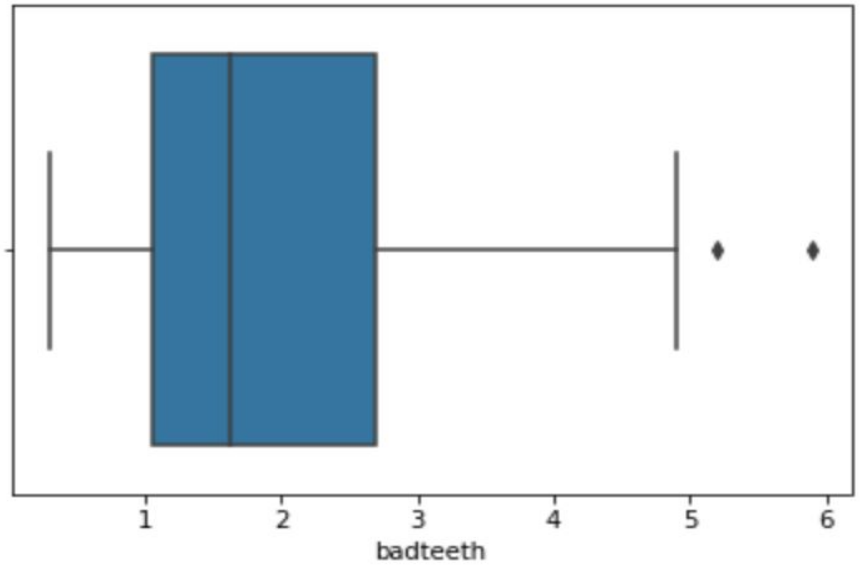
Introduction to World Dental Risks

The WHO website claims risk factors for oral diseases include an unhealthy diet, tobacco use and alcohol use. Other risk factors include a high sugar diet, poor hygienic habits, and lack of professional care.

High sugar diets, tobacco and alcohol use are linked with a higher GDP, where the population can afford these unhealthy luxuries. On the other hand, these countries are more likely to have better health education and more access to professional dental care.

Bad Teeth Overview

“Bad Teeth” is the number of teeth with some sort of dental issue for the average child aged 12 and under for that country. The studied countries have numbers ranging from 0.3 to 5.9 teeth with problems with a mean of 1.98 and median of 1.625.



Data Used

- **Badteeth** - Bad teeth per child under 12 years old, by country
- **Literacy_avg** - Percent of adults age 15+ who are literate, by country
- **Water_avg** - Percent of population with access to basic drinking water sanitation by country
- **GDP_avg** - GDP per capita in US dollars and inflation adjusted, by country
- **Sugar_avg** - Sugar consumption per person in grams per day by country
- **Health_avg** - Government health spending per person by US dollars by country
- **Low_BMI** - Percent of youth > 2 standard deviations below median world BMI by country
- **Adolescent birth rate** - Adolescent birth rate per 1000 women age 15-19 by country
- **Tobacco_use** - Percentage tobacco use by youth aged 17 or younger by country
- **Smokers** - Percent smoking rate by youth, aged 17 or younger by country

This data is from 2004 or before and does not fully represent current data

Goals and Tools Used

The main goal of this project is to explore links between basic demographic data and dental health and to group similar countries together based on demographic data and find a trend in the number of childhood dental problems.

The tools used include:

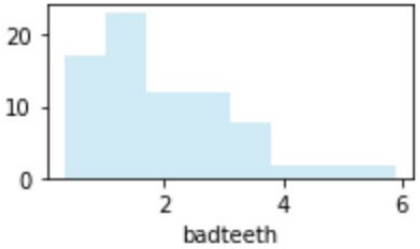
Pandas - to read and save files, convert types,

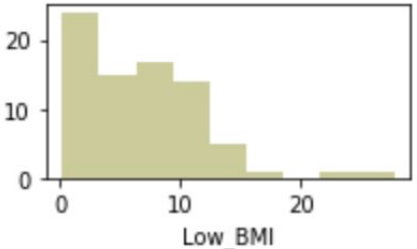
Seaborn, matplotlib - for visualizations

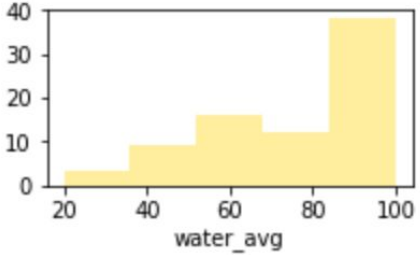
Sklearn - Machine Learning, linear regression, clustering, PCA

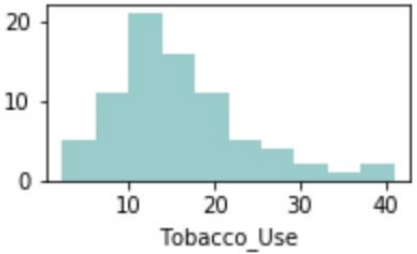
Scipy - for correlations, significance tests, dendrograms, linkage, fcluster

A Look at the Demographics



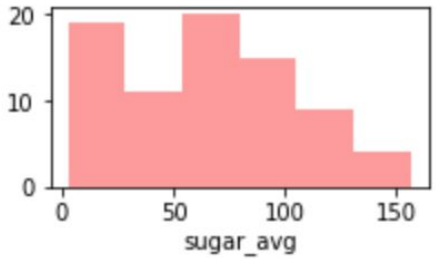


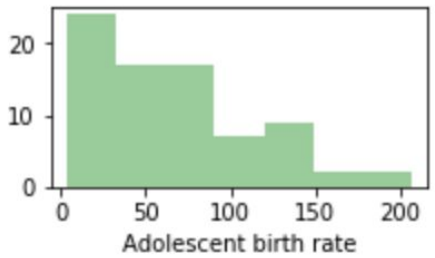


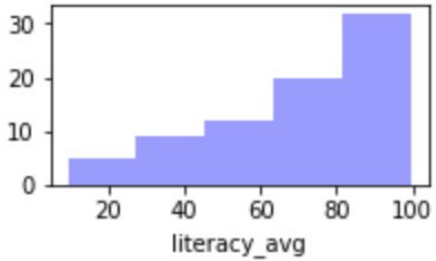


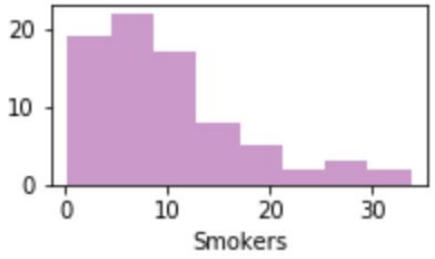
Half of the countries in our study had more than ¾ of their population with access to basic sanitized drinking water. The highest percent of children with severely low BMI was 27.7% (India), but more than ¾ of countries' have less than 10 percent. Most tobacco use hovers around 15 percent of the population.

Demographics Cont.



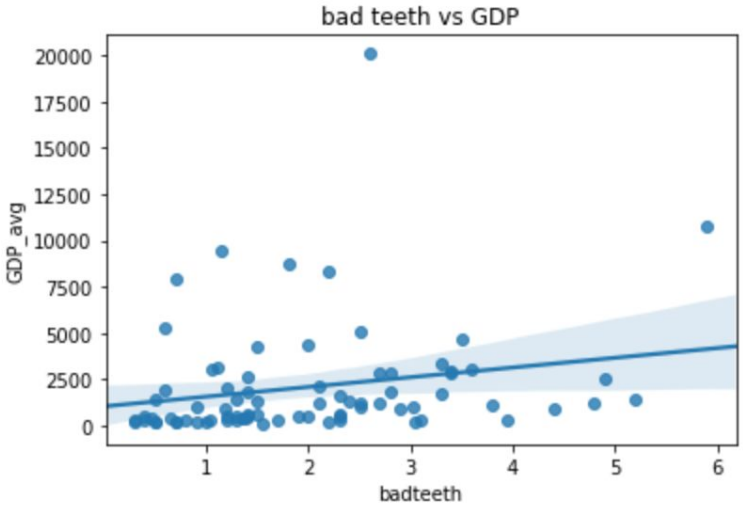


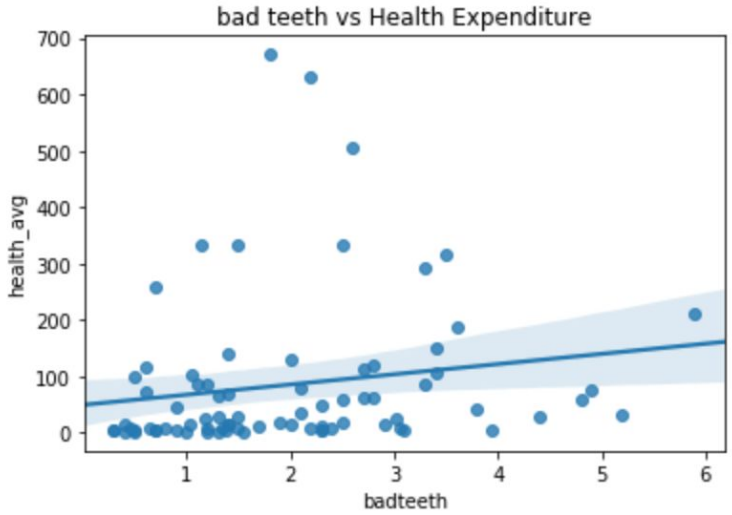




The distribution of countries' literacy rates are heavily skewed left, while sugar consumption has an almost normal distribution. Both adolescent birth rates and smoking rates are mostly very low. Samoa and Latvia hold the highest youth smoking and tobacco use rates. The highest 17 adolescent birth rates by country are all located in Africa.

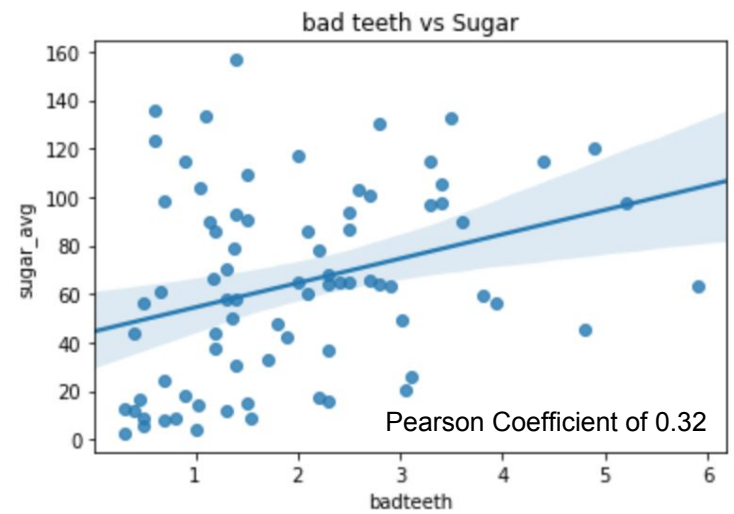
Is Money Helping?

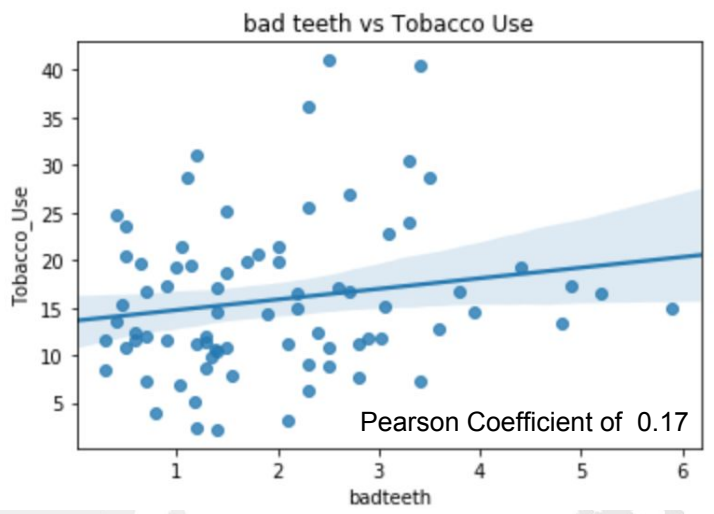




While it requires a higher GDP to have a higher healthcare budget, there is a very slight positive correlation (Pearson Coefficient of 21.4) between dental issues and GDP, meaning money does not help, possibly because more money is linked to higher sugar consumption.

How do Sugar and Tobacco Come in?





Both Sugar consumption rate and tobacco use have a positive correlation with dental problems in children. Sugar most likely has a stronger correlation because tobacco is less often used by children under 12 and its effects are not immediate.

Countries that Stand Out

While as a whole, African countries have a lower GDP, higher adolescent birth rate and children tend to have a lower BMI, African children tend to have less dental health problems. Countries with high rates of problem teeth in children are scattered across the globe.

	Countries	badteeth
58	Rwanda	0.30
68	Togo	0.30
36	Lesotho	0.40
37	Liberia	0.40
52	Nigeria	0.46

	Countries	badteeth
73	Ukraine	4.4
7	Bosnia and Herzegovina	4.8
43	Mauritius	4.9
25	Guatemala	5.2
60	Saudi Arabia	5.9

Can We Group the Countries? Why Should We?

The goal here is to try to group similar countries together to create similar plans for improvement and aid. If we can group similar countries together, they should react to aid similarly and we can use a success in one country to help other countries in the same group.

Before we can group countries though, the data needs to be cleaned and prepped. Missing values need to be dealt with and all extraneous data removed.

Scaling

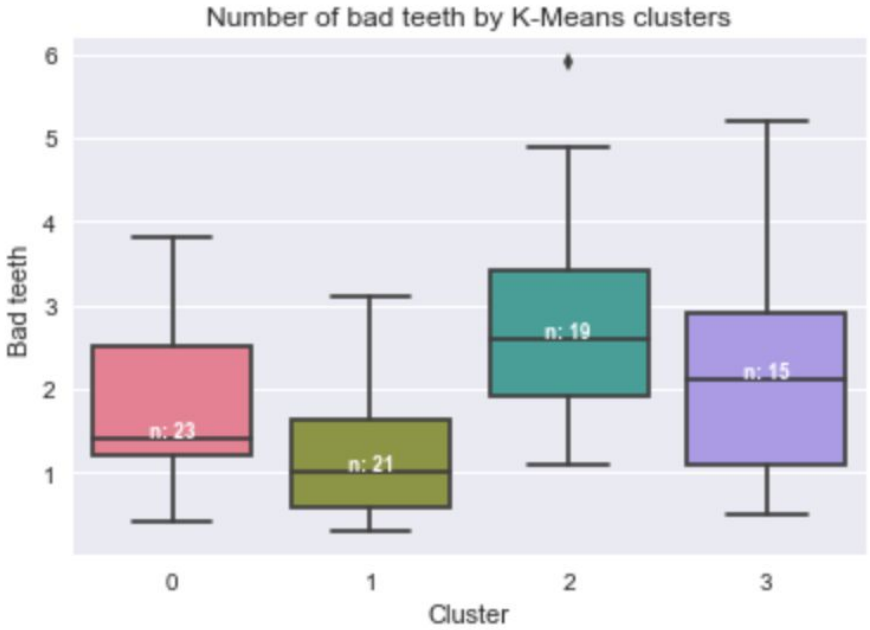
Scaling is a very important process that puts all variables on the same or very similar scale to make more accurate comparisons.

Using Min-Max Scaling, the data is scaled to a fixed range between 0 to 1.
This creates smaller standard deviations and lessens the effects of outliers. Scaling the data is important because the model is sensitive to magnitude, and the units of each feature is different.

K Means Clustering

The first clustering method I used is K-means clustering. K-Means clustering works by randomly selecting k number of points in your data to be your temporary centers. The clusters are then created by assigning each data point to the nearest center. The algorithm evaluates a new center of this created cluster and the process is repeated until it reaches convergence. In this case I had it split the countries into four groups. This algorithm then clusters the countries based on all 9 variables of the scaled data (not including the number of bad teeth)

K Means Results



n is the size of the cluster

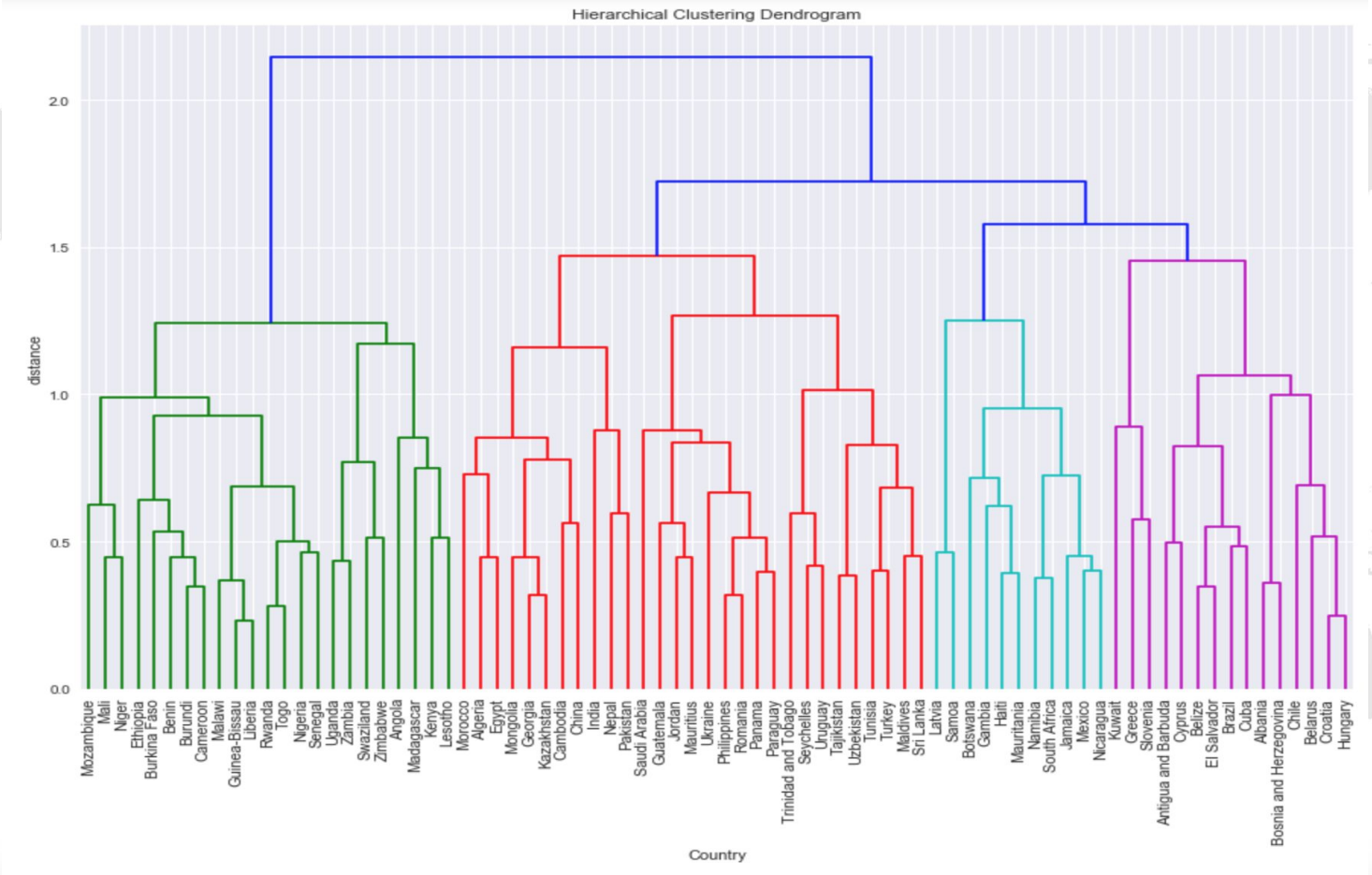
It looks like the clusters do have different ranges of bad teeth. Cluster 1 holds the countries with the lowest number of bad teeth, then cluster 1, 3, then 2 has the most countries with higher rate of dental issues. Cluster 3 has a much wider range than the others, yet the least amount of countries in it. The diamond signifies an outlier, which is Saudi Arabia.

Hierarchical Clustering

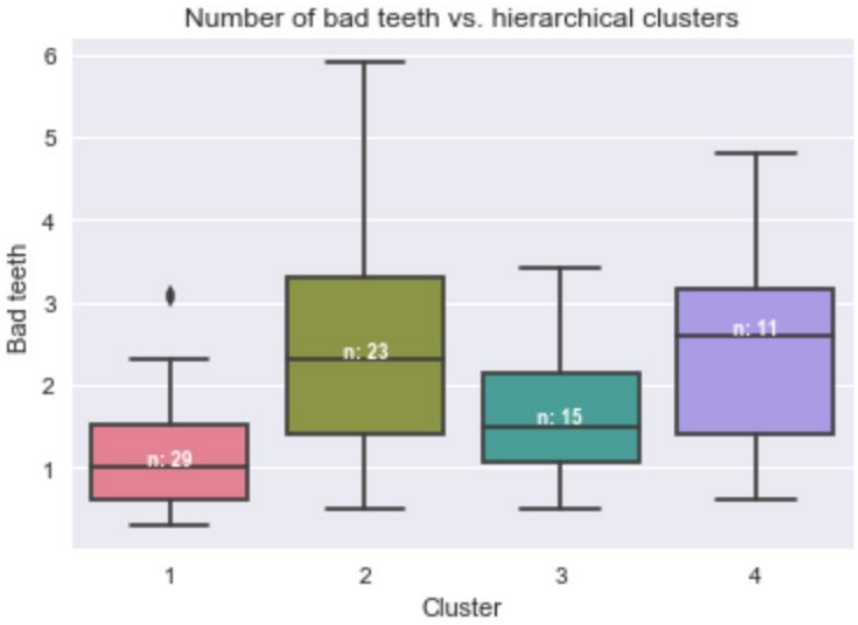
The complete linkage hierarchical clustering works by starting with each element as its own cluster. In this case, each Country starts as a cluster, then the closest clusters get combined. These new clusters then are merged with the closest clusters and so on. What makes it a “complete” linkage is that the distances between the clusters are measured by the furthest two points in the clusters. You decide how many clusters to choose by changing the distance at which you stop at.

A dendrogram is a great way to visualize this. The y-axis of a dendrogram is the distance between each cluster, while the countries are along the x-axis.

Dendrogram for Hierarchical Clustering



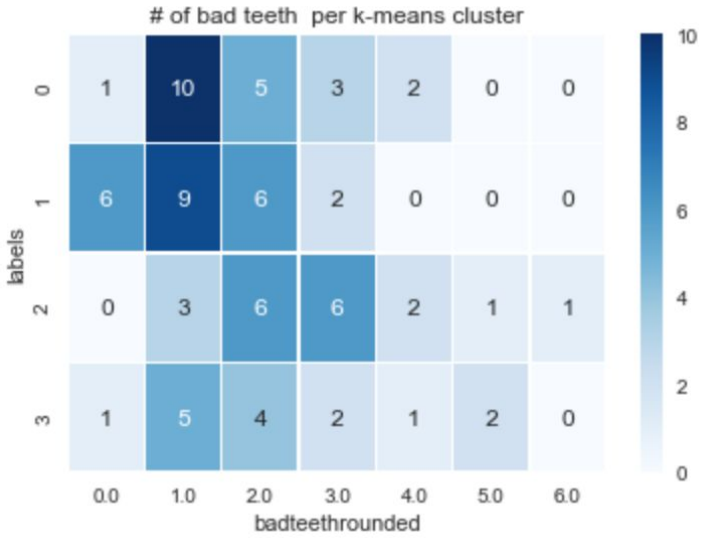
Hierarchical Clustering Results

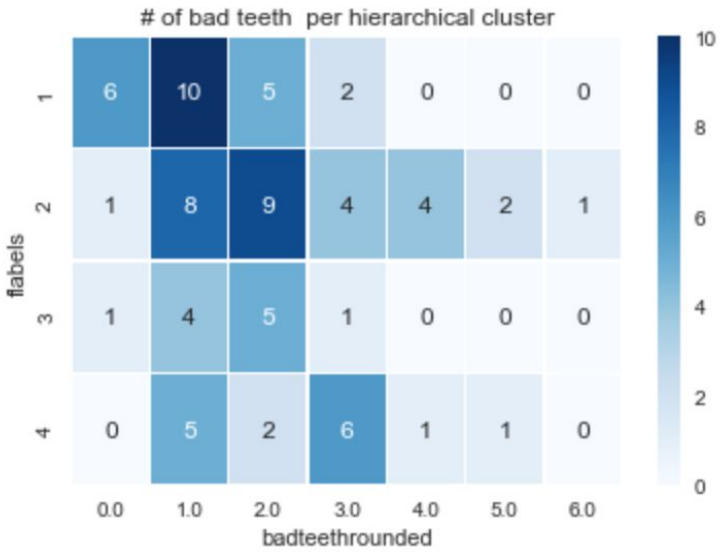


n is the size of the cluster

While they are in a different order, the hierarchical cluster distributions look somewhat similar. One cluster has very low numbers, then two middle clusters, and a final with a very large range. This clustering result differs in that the very low cluster has the highest number of clusters and the smallest cluster has only 11 countries.

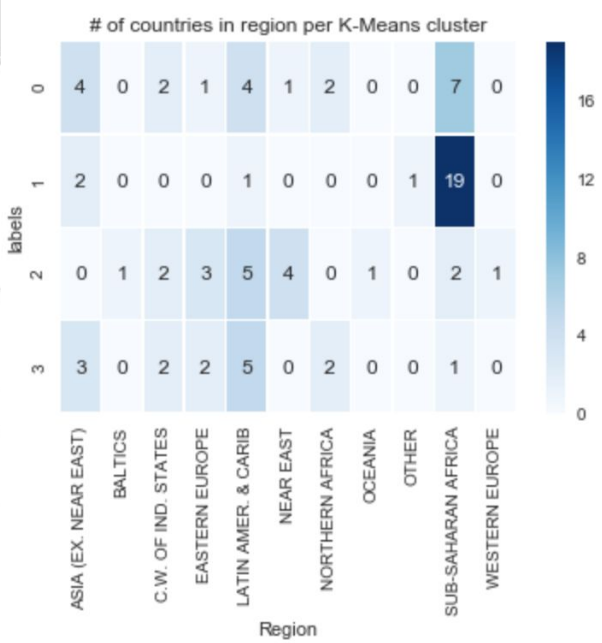
Comparing Methods and Dental Health



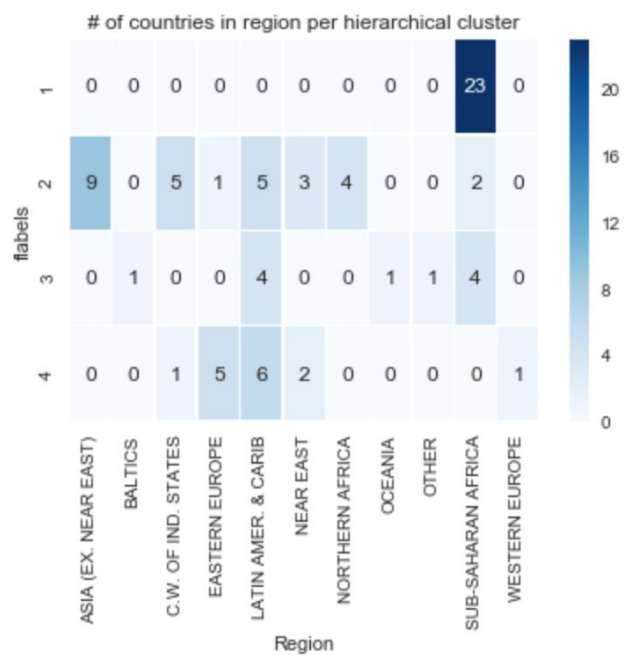


Both methods gave very similar results for cluster 1 in K-means and cluster 1 in the hierarchical. The other clusters seem to differ a bit more, and cluster 4 of the hierarchical has a bimodal distribution

Comparing Regions and Dental Health



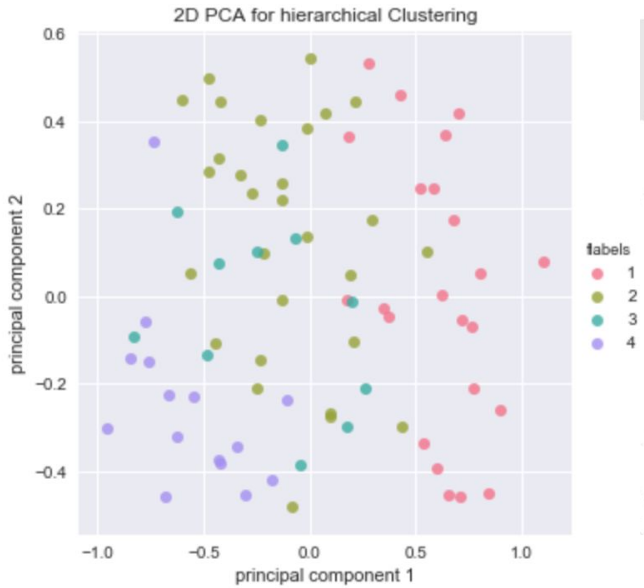
It looks like cluster 1 is almost entirely made up of sub-saharan countries for both methods and the northern African countries are split between cluster 0 and 3 for K-means, and all the northern African countries are in the same cluster for the hierarchical clustering method. Other than this there does not seem to be any obvious trends between cluster and region.



Visualizing Clusters

Dendrograms are a great way to visualize hierarchical clustering, but other forms of clustering have less options for visualizing so many dimensions. PCA, or Principal Component Analysis is used in this case to reduce the number of dimensions from 10 variables to 2 so we can plot the clusters on a 2d graph. I chose not to use the PCA information for the clustering because I did not have a very large amount of data, and after finding the cumulative explained variance ratio I found that reducing to 2 components would retain only 57.7% of the variance from the original data.

PCAs



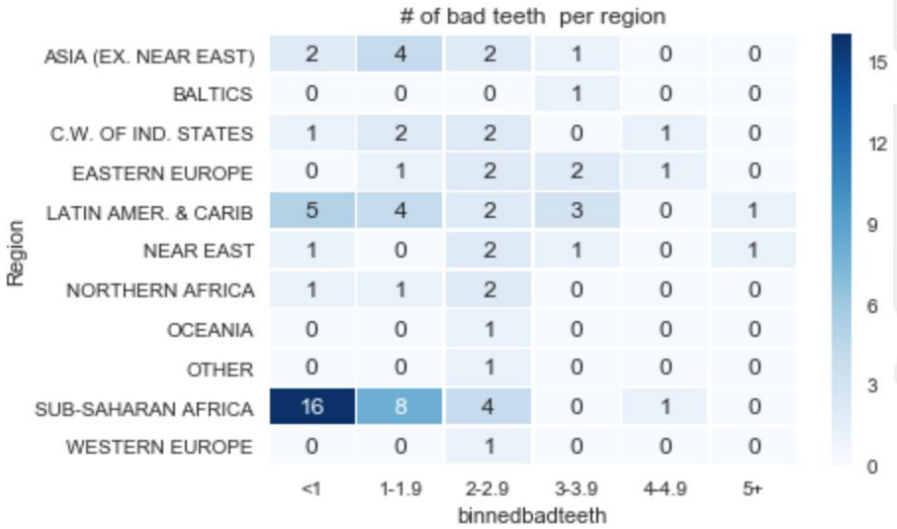
If the PCA retained more than the 60% variation, I would look at these graphs and say K-means did a better and more precise clustering. With what we have, this really cannot be used to assert those kinds of claims

Method Summaries

The Hierarchical clustering seemed to cluster the extremes of the bad teeth very well. The ten countries with the largest amount of childhood tooth problems were all grouped into 2 clusters and the lowest six countries were all in the same cluster. The K-means method's clusters were more consistent in size and had a cleaner PCA. Either method would be reasonable to base a plan off of.

Regions vs. number of bad teeth

Could we simply group the countries by region instead of by cluster?



While you could group the countries by region and lump the countries that way, this does not make up for the outliers and extremes. Overall it would do an OK job, but not as accurate as either clustering method. You would also have 4 clusters with only one country in it.

Conclusions and Trends

Sub-saharan Africa was very distinct from the other regions and the countries were quite similar to one another. The seven countries with the least bad teeth are all from Sub-saharan Africa. This is not without exceptions though, as Mauritius has the third highest child dental issues. In general, wealthier countries with more access to clean water and higher literacy rates and overall health had experienced slightly more teeth issues in children.

The World Health Organization website claims that risk factors for oral diseases include unhealthy diet, tobacco use and alcohol use. All of these are easier to access with a wealthier population. I am going to suggest, given the data, that social determinants and general culture around dental hygiene explains a large portion of the variance of oral health in children. Countries with 5 times the amount of health expenditure per person have the same if not more oral issues than other countries, and no variable has a remarkable correlation. Culture being a strong factor would also explain why some regions are so different from others.

Limitations and Further Research

Given the low amount of data, there was some struggle with finding useful results. Further analysis would benefit from more data, which could come in the form of more countries included, data from more recent years, and/or more variables to compare. I would also love to see how much different dental procedures cost in each region to compare a return of investment or compare the availability of current oral health services. Additionally, the WHO claims that fluoride in drinking water can significantly reduce cavities in the entire population. I would recommend doing more research or conducting a survey to see which countries currently have flouride in their public drinking water and how much of the population has access to it. I would love to see if there is actually a correlation in the data. If possible it might even be beneficial to break down the country further into sections as to reduce the misrepresentation of a country with a large wealth gap.